面向低资源命名实体识别的 BiLSTM-BCRF 模型[1]

钟茂生 吴佳华
江西师范大学计算机信息工程学院

摘要：[目的]当标注数据较少时，现有模型受训练数据量少的限制，参数没有拟合到预期效果，导致在低资源命名实体识别任务中模型识别性能不佳。[方法]本文提出一种融入伯努利分布(Bernoulli distribution)的新型损失函数，让模型较好拟合数据。此外，本文在 BiLSTM-CRF 模型基础上融合多层字符特征信息，结合基于伯努利分布的新型损失函数，构建了 BiLSTM-BCRF 模型。[结果]本文提出的 BiLSTM-BCRF 模型在 20%的 CoNLL2003 和 20%的 BC5CDR 的数据集上，F1 值在 BiLSTM-CRF 模型基础上分别提升了 6.16%、3.35%。[结论]该模型能较好地适应低资源命名实体识别任务。[局限]该模型识别专有名词的性能还有待提升。

关键词：低资源命名实体识别; 神经网络;伯努利分布

BiLSTM-BCRF Model for Low Resource Named Entity Recognition

Zhong Maosheng, Wu Jiahua
School of computer information engineering, Jiangxi Normal University

Abstract: [Objective]when there are few labeled data, the existing models are limited by the amount of training data, and the parameters do not fit the expected effect, resulting in poor model recognition performance in the task of low resource named entity recognition. [Methods] a new loss function integrated with Bernoulli distribution is proposed to make the model fit the data better. In addition, based on the BiLSTM-CRF model, multi-layer character feature information is fused, and the new loss function based on Bernoulli distribution is combined to construct the BiLSTM-BCRF model. [Results] Based on the dataset of 20% CONLL2003 and 20% BC5CDR, the F1 value of the BiLSTM-BCRF model proposed in this paper increased by 6.16% and 3.35% respectively. [Conclusion] the model can better adapt to the task of low resource named entity recognition. [Limitations] the performance of this model in identifying proper nouns needs to be improved

Keywords: Low resource named entity recognition；Neural network；Bernoulli distribution

1 引言

命名实体识别是自然语言处理的基础任务之一，该任务旨在从非结构化的文本中自动识别出实体，并将其标记为预定义的类别，例如人名、地名和组织机构名等。例如，“张无忌，金庸武侠小说《倚天屠龙记》人物角色，中土明教第三十四代教主。”这句话包含的实体有：人名实体“张无忌，金庸”，书名实体“倚天屠龙记”，门派实体“明教”。由此可见,实体识别是文本语义理解的基础。同时命名实体识别技术在知识图谱构建、机器翻译、知识库构建等多种自然语言处理任务中有着广泛应用。

近些年来，深度学习方法被广泛用于命名实体识别，如 Hammerton[1]将长短期记忆网络(LSTM)应用到实体识别研究中，LSTM-CRF 结构成为实体识别的基础结构。Lample 等人[2]在 LSTM-CRF 模型的基础上，提出双向长短期记忆网络(Bi-LSTM)和条件随机场(CRF)结合的模型[3-5]等。这类方法虽然在文本实体识别任务中表现优异 ，但是需要大规模的标注数据，对训练语料中每个词进行人工标注。在标注数据不足的情况下，现有模型的参数不能较好拟合，导致模型预测最大概率标签并不一定是真实标签，模型的识别性能下降，很难应用到如生物、医学这些标注语料较少的领域。针对上述问题，本文提出一种融入伯努利分布的新型损失函数,使模型参数在

1 作者：吴佳华，ORCID：0000-0002-5704-8406，E-mail:202041600071@jxnu.edu.cn;
基金：本文系国家自然科学基金项目“面向在线智慧学习的多模态学习资源组织与个性化推荐服务研究 ”（项目编号：NO.61877031）的研究成果之一。

低资源场景下能较好拟合。在此基础上，为增加模型能处理的词汇量和提升模型识别罕见词的能力，本文在 BiLSTM-CRF 模型基础上融合多层字符特征信息,进一步提升了模型的精确率和召回率。

本文的组织结构为：第二节介绍低资源命名实体识别领域的主要工作。第三节介绍本文模型，包括：输入层、BiLSTM 层和 BCRF 层。第四节介绍实验数据、实验内容、实验结果及分析。最后对本文工作进行总结。

2 相关工作

命名实体识别的研究方法主要有基于规则和词典方法、机器学习方法、深度学习方法等。基于词典和规则的方法过多依赖于语言学家制定的规则模板，容易产生错误，移植性差。传统机器学习方法主要包括:隐马尔可夫模型(Hidden Markov Model,HMM)、最大熵(Maximum Entropy,ME)[6]、最大熵马尔可夫模型(Maximum Entropy Markov Model,MEMM)[7]、条件随机场(Conditional Random Fields,CRF)[8]等。这些传统的机器学习方法在特征提取方面需要人工参与，同时需要大规模的标注语料来训练模型，方法的性能主要依赖于所采用的特征是否具有辨识度。其中 CRF 被看作是命名实体识别的主流模型,优点在于在对一个位置进行标注的过程中 CRF 可以利用内部及上下文特征信息。随着深度学习的不断发展,命名实体识别的研究重点已转向深层神经网络, Collobert 等学者[9]首次提出基于神经网络的命名实体识别方法,该方法中每个单词具有固定大小的窗口,但未能考虑长尾问题。为了克服这一限制,Chiu 和 Nichols[10]提出一种双向 LSTM-CNNs 架构,该架构可自动检测单词和字符级别的特征。Hammerton[1]利用 CRF 关注上下文特征信息的特点，提出 LSTM-CRF 模型。

近些年来,大量的深度学习方法被应用于低资源命名实体识别任务中，低资源的命名实体识别技术成为当前研究热点之一，其性能的提高是命名实体识别技术走向广泛实际应用的前提。相关研究工作可大致分为以下几类： 跨语言迁移的方法、数据增强的方法、集成自动标注语料的方法和其他方法。

跨语言迁移方法的基本思路是利用资源丰富语言的标注数据帮助低资源语言进行命名实体识别,可大致分为数据迁移的方法和模型迁移的方法两大类。基于数据迁移的方法通常借助文本翻译和标签映射等手段把源语言中的标注数据转换成目标语言的标注数据，然后基于这些数据训练模型。Ni 等[11]提出了一种在语料库上进行标签映射的方法，用于创建自动标记的目标语言数据。Mayhew 等[12]利用双语词典，使用一种类似短语机器翻译[13]的方法自动翻译源语言的标注文本。基于模型迁移的方法通常先学习语言无关的特征，然后在源语言的标注语料上训练 NER 模型直接用于目标语言。Chen 等[14]同样基于对抗学习的方法提取语言无关的特征，并动态地计算源语言和目标语言之间的相似度，从而更有效地实现从多个源语言到目标语言的知识迁移。Keung 等[15]在多语言版本 BERT 的基础上进一步使用对抗学习[16]的方法，以学习更好的与语言无关的特征。

数据增强方法的主要目标是,在不增加人工标注成本的前提下,通过增加合理的噪声来提升模型的鲁棒性， 在少数据量的场景下对模型性能的提升有很大帮助。Dai 等[17]引入了一些词替换的随机操作来增加训练语料多样性;Chen 等[18]在半监督 NER 任务中引入了基于局部可加性的数据增强。基于语言迁移的方法和数据增强的方法虽然能够有效地缓解标注语料短缺的问题，但是具有丰富标注资源的语言是非常少的。为此，一些研究者提出集成自动标注语料的方法，首先通过某种方法自动标注大量语料，然后集成它们用于提高低资源实体识别模型的性能。Yang 等[19]首先基于词典匹配的方法自动标注语料，然后使用 Partial-CRF[20]在少量人工标注的语料和大量自动标注的语料上训练实体识别模型。此外，他们还基于强化学习[21]训练一个选择器，用于筛选掉具有噪声的标注数据。

除上述三类方法外，低资源实体识别领域还有其他方法如 Zhang 提出渐进式知识提炼方法 PDALN[22]，有效的将高资源域适应于低资源目标域。Chen 提出了一种低资源的语言模型的微调方法[23]，使用基于注意力机制的微调策略，从预训练的语言模型中选择相关的语义和句法信息，将其应用于命名实体识别任务。本文的工作主要是探索低资源条件下基于深度学习的命名实体识别方法。

3 模型

3.1 基本架构

命名实体识别任务被看作是序列标注问题。输入句子表示为,其中表示第 i 个字符(包括数字、单词、字母或标点符号等)。输出标注序列为,其中∈{B,M,E,S,O}是的标签,B、M、E、S 和 O 分别代表实体首字,实体中间字,实体结尾字,实体单独字和非实体。命名实体识别就是对每个字符进行 B、M、E、S、O 的分类标注。

本文在 BiLSTM-CRF 模型损失函数中融入伯努利分布，同时将多层字符信息融合到 BiLSTM-CRF 模型，构建了 BiLSTM-BCRF 模型。BiLSTM-BCRF 模型基本结构如图 1 所示。该模型结构主要分为输入层、Bi-LSTM 层和 BCRF 层。

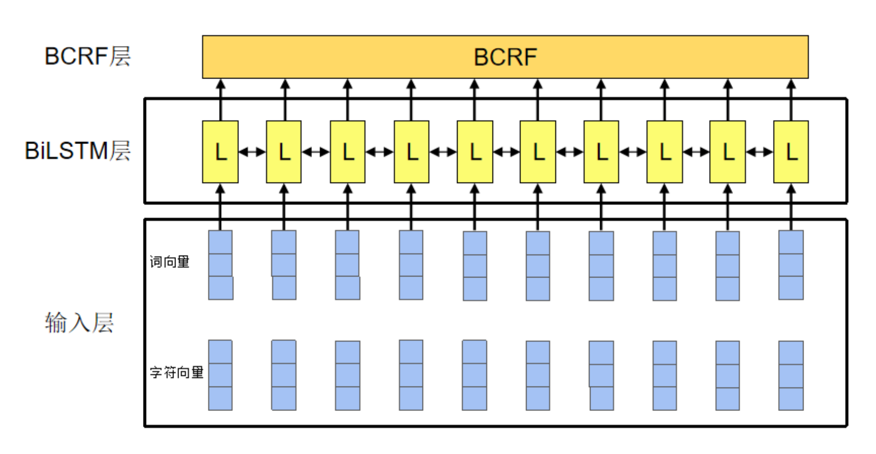


图 1 BiLSTM-BCRF 模型基本结构

Figure 1 basic structure of BiLSTM-BCRF model

3.2 输入层

如图 2 所示，该图为模型输入层的结构图。其中，x 表示词向量，是使用 Pennington 等人提出 Glove 英文词向量[24]文件生成的，；表示由 BiLSTM 训练生成的字符向量，。最后将词向量与字符向量拼接输入到 BiLSTM 层。

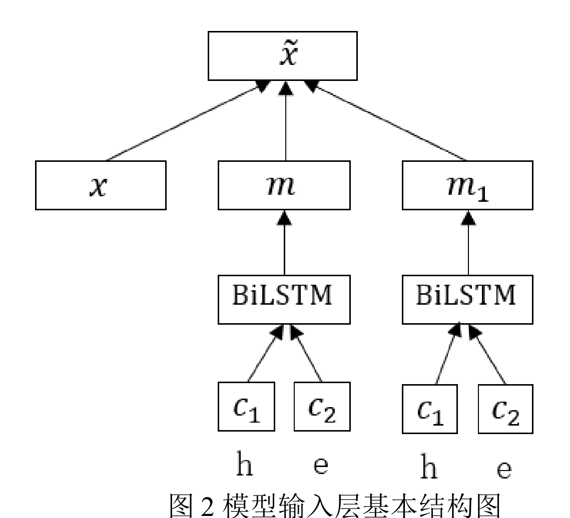


图 2 模型输入层基本结构图

Fig. 2 basic structure of model input layer

3.3 BiLSTM 层

LSTM 神经网络在命名实体识别任务中表现出良好的建模能力，能较好的学习文本中单词与字符的特征信息，BiLSTM 层结构主要由两个 LSTM 组合而成。LSTM 的网络结构主要分三个阶段：遗忘阶段、选择记忆阶段、输出阶段。LSTM 单元结构如图 3 所示，分别表示 LSTM 单元中

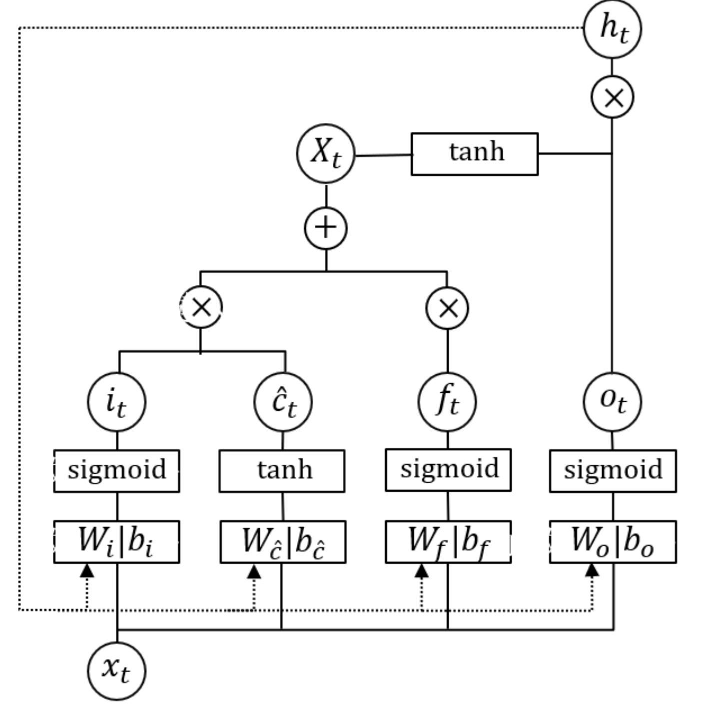


图 3 LSTM 单元结构图
Figure 3 structure diagram of LSTM unit

的输入门、遗忘门、输出门在 t 时刻的状图态。表示在 t-1 时刻的隐藏状态,表示在 t 时刻的细胞记忆状态。σ 表示 Sigmoid, tanh 表示双曲正切激励函数, 如式(1)～式(6)所示：

(1
)

(2)

(3)

(4)

(5)

(6)

BiLSTM 神经网络中的输出隐藏状态,分别为前向输出和后向输出。

3.4 BCRF 层

原有命名实体识别模型解码层主要是条件随机场(Conditional Random Field，CRF)。CRF 是由状态特征函数和状态特征转移函数组成，状态特征函数也称发射概率，状态特征转移函数在模型中可以用一个状态转移矩阵表示，最后得到的条件概率如式(7)所示：

(7)

其中，是 x 和 y 一组特征向量的映射。表示模型在给定文本序列 x 条件下得到标签序列 y 的概率。损失函数计算公式如式(8)所示：

(8)

CRF 方法的优点是可以进一步考虑序列标签的依赖关系，同时在训练过程中，采用 Viterbi 算法用于最大似然估计，使模型对输入文本预测出标签的最大概率如式(9)所示：

(9)

其中，表示模型预测标签的最大概率。但是在低资源的场景下，模型受标注数据量少的限制，参数没有拟合到预期效果，输出预测概率最大的标签序列并不一定是真实的标签序列，导致模型最后的识别性能下降。在此，本文借鉴 jie 等人[25]在不完全标注实体识别任务上采用交叉验证方式训练数据的思想，将伯努利分布融入到条件随机场损失函数中，构建新的损失函数，对应的解码模型称为 BCRF。

伯努利分布(Bernoulli distribution)又名两点分布或 0-1 分布，它是指对于随机变量 X，参数为 P(0<P<1),它分别以概率 P 和 1-P 取 1 和 0 为值，该分布是一个离散型概率分布，是 N=1 时二项分布的特使情况。将伯努利分布函数融入到 CRF 中，构建新的损失函数如式(10)所示：

(10)

在原有的损失函数计算公式加入一个分布函数 q，计算公式如式(11)所示。分布函数 q 取值为 0 或 1，呈伯努利分布。

(11)

其中，表示模型预测最大概率的标签，表示真实标签。在式(10)中，当句子中词预测标签与真实标签一致时，得到的损失值也就较小。那么相反，如果在一个句子中预测标签错误的越多，那么它的损失值也越大。采用新的损失函数公式计算，在训练样本只有少量时，模型参数能较好拟合，提升模型的精确率和召回率，最后提高模型的识别效果。

4 实验

4.1 数据集及评价指标

本文选择 CONLL2003[26]数据集和 BC5CDR[27]数据集来证明所提出模型的有效性。CONLL2003 数据集包含 4 种实体类型以及英语和德语两种语言，BC5CDR 数据集包含两种实体和 1500 篇医药文章。由于实体识别的任务主要是对实体的边界和类别的识别，只有当边界及实体的类别都识别正确时，才判断正确。本文通过使用精确率(Precision)和召回率(Recall)来求得 F1 值，用于衡量该模型的性能，如式(12)式(14)所示：

(12)

(13)

(14)

其中 N 表示为模型所预测出的实体总数，M 表示模型预测的实体中正确预测实体的总数，K 表示为数据集中所标注的实体总数。模型中超参数设置如表 1 所示：

表 1 超参数设置

Table 1 super parameter setting

参数	值
隐层向量维度	200
词向量维度	100
字符向量维度	50
Dropout	0.5
学习率	0.1
批尺寸	10
训练轮数	100
L2 正则化	1e-8

4.2 实验结果与分析

本文所做的实验采用的数据集主要是 CONLL-2003 英语数据集和 BC5CDR 专业医学领域数据集，BiLSTM-BCRF 模型与 BiLSTM-CRF 模型实验结果如图 4、图 5 所示：

图 4 BC5CDR 数据集上 BiLSTM-BCRF 模型与 BiLSTM-CRF 模型实验结果对比

Fig. 4 Comparison of experimental results between BiLSTM-BCRF model and BiLSTM-CRF model on BC5CDR dataset

图 5 CONLL2003 数据集上 BiLSTM-BCRF 模型与 BiLSTM-CRF 模型实验结果对比

Fig. 5 Comparison of experimental results between BiLSTM-BCRF model and BiLSTM-CRF model on CONLL2003 dataset

从图 4、图 5 中可以看出该模型在少量标注数据集上的性能是高于 BiLSTM-CRF 模型，比较适合低资源领域的实体识别任务。同时该模型在 30%的 CONLL2003 数据集上 F1 值达到了 89.32%，说明该模型在不需要大量的标注语料的情况下，也能取得比较好的识别效果。

TMN 是 Lin 等人[28]在 2020 年提出的一种基于实体和触发词标注的命名实体识别模型。BiLSTM-BCRF 模型与 TMN 模型实验结果对比如表 2、表 3 所示。

表 2 BiLSTM-BCRF 模型与 TMN 模型实验结果对比

Table 2 Comparison of experimental results between BiLSTM-BCRF model and TMN model

20%CONLL2003			
	Precision	Recall	F1
BiLSTM-CRF	82.17	80.35	81.3
TMN	86.04	85.98	86.01
BiLSTM-BCRF	87.49	88.63	**88.06**

如表 2 所示，该实验是在 20%的 CONLL-2003 数据集上进行的。从表中可以看出 BiLSTM-BCRF 模型 F1 值是是高于 TMN 模型的。

表 3 BiLSTM-BCRF 模型与 TMN 模型实验结果对比

Table 3 Comparison of experimental results between BiLSTM-BCRF model and TMN model

20%BC5CDR			
	Precision	Recall	F1
BiLSTM-CRF	79.09	62.66	69.92
TMN	77.47	70.47	**73.97**
BiLSTM-BCRF	74.35	72.22	73.27

如表 3 所示，在 20%的 BC5CDR 数据集上，BiLSTM-BCRF 模型 F1 值比 BiLSTM-CRF 模型高 3.35%，比 TMN 模型低 0.7%，主要是该模型识别一些专有名词的性能略低于 TMN 模型。但是，BiLSTM-BCRF 模型并不需要标注触发词，它所需要的人工成本只是 TMN 模型的 3/4，同时 BiLSTM-BCRF 模型在 CONLL2003 数据集上 F1 值比 TMN 模型高 2.05%。当然，如何提高 BiLSTM-BCRF 模型识别专有名词的性能，也是本文后续工作的重点。

4.3 词与多层字符信息的融合分析

在 3.2 节中本文提出如图 2 所示的模型输入层结构, 为验证在模型输入层哪些因素会影响模型进行命名实体识别的性能，本文进行如下实验：

表 4 词向量与字符向量拼接顺序对模型性能的影响

Table 4 explores the impact of word vector and character vector splicing order on model performance

20%BC5CDR 数据集			
Method	Precision	Recall	F1
word+char	71.82	72.50	72.16
char+word	71.23	73.07	72.14
BiLSTM-BCRF (word+char*2)	74.35	72.22	**73.27**
char+word+char	72.25	71.73	71.99

表 4 中的实验所使用的模型为 BiLSTM-BCRF 模型，采用的数据集为 20%的 BC5CDR 数据集。其中，word 表示为词向量信息，维度设置为 100。char 表示为字符向量，维度设置为 50；“char*2”表示为 2 个字符向量矩阵拼接，“+”表示拼接。

为探究词向量与字符向量拼接顺序是否会影响模型的性能，本文做了以下实验，实验结果如表 4 所示：在模型输入层采用词向量和一个字符向量矩阵拼接时，将它们的拼接顺序调换，模型 F1 值与之前相比，稍微下降。当词向量与两个字符向量矩阵拼接时，将词向量放至两个字符向量矩阵中间时，模型的精确率、召回率、F1 值与之前相比都有所下降。由此可知，词向量与字符向量的拼接顺序是会影响模型的性能的

表 5 词向量拼接字符向量矩阵的个数对模型性能的影响

Table 5 explores the impact of the number of word vector stitching character vector matrices on the performance of the model

20%BC5CDR 数据集			
Method	Precision	Recall	F1
BiLSTM-BCRF (word+char*2)	74.35	72.22	**73.27**
word(200)+char*4	73.12	73.20	73.16
word(300)+char*6	70.74	73.87	72.27
word(50)+char	70.42	70.70	70.56
word+char	71.82	72.50	72.16
word+char*3	71.07	74.21	72.60
(word+char*2)*2	73.48	71.37	72.41

表 5 中的实验所使用的模型为 BiLSTM-BCRF 模型，采用的数据集为 20%的 BC5CDR 数据集。其中，word 表示为词向量信息，维度设置为 100。“word(200)”表示维度为 200 的词向量，char 表示为字符向量，维度设置为 50；“char*3”表示为 3 个字符向量矩阵拼接，“+”表示拼接。

在低资源场景下，模型受标注数据量少的限制，通过在词向量后拼接字符向量，可以提高模型处理罕见词的能力，提高模型的识别性能。为探究拼接字符向量矩阵数量为多少时，模型识别性能提升的最多，本文做了以下实验，实验结果如表 5 所示：首先将词向量维度设置为 100 时，

拼接一个字符向量矩阵、两个字符向量矩阵、三个字符向量时矩阵，通过实验结果对比，本文发现在词向量后拼接两个字符向量矩阵，模型识别效果最好。然后将词向量维度设置为50、200、300时，后拼接不同数量的字符向量矩阵。通过实验结果对比，本文发现将词向量维度设置为100，拼接两个字符向量矩阵，模型的识别性能提升的最多。

4.4 消融实验

为探究伯努利分布和多层字符信息对模型性能的影响,本文将 BiLSTM-CRF 模型设置为基准模型，首先将伯努利分布融入到基准模型的损失函数中，构建 BiLSTM-BCRF(1)模型。然后在 BiLSTM-CRF(1)模型基础上融合多层字符信息，构建 BiLSTM-BCRF 模型，以上模型在两个数据集的实验结果如表 6 所示：

表 6 消融实验结果

(BiLSTM-CRF 模型为基准模型，BiLSTM-BCRF(1)模型表示将伯努利分布融入到 BiLSTM-CRF 模型损失函数中)

Table 6 ablation experimental results

(the BiLSTM-CRF model is the benchmark model, and the BiLSTM-BCRF (1) model represents the integration of Bernoulli distribution into the loss function of the BiLSTM-CRF model)

	20%CONLL2003			20%BC5CDR		
	Precision	Recall	F1	Precision	Recall	F1
BiLSTM-CRF	82.17	80.35	81.3	79.09	62.66	69.92
BiLSTM-BCRF(1)	86.43	87.75	87.08	73.77	71.98	72.87
BiLSTM-BCRF	87.49	88.83	**88.06**	74.35	72.22	**73.27**

表 6 所示的消融实验结果表明，BiLSTM-BCRF(1)模型在两个数据集上取得的 F1 值均高于 BiLSTM-CRF 模型，表明新型损失函数对于模型的性能具有提升的作用。从表中模型 BiLSTM-BCRF 在两个数据集取得的精确率、召回率、F1 值可以看出，本文提出的 BiLSTM-BCRF 模型识别效果是比较好的。

4.5 定性分析

为更好的对比 BiLSTM-BCRF 模型与 BiLSTM-CRF 模型在命名实体识别任务上的差异，本文从数据集中选取两个实例句子：“Only France and Britain backed Fischler ‘s proposal.”和“Rare Hendrix song draft sells for almost $ 17000”，人工标注、BiLSTM-CRF 模型和 BiLSTM-BCRF 模型的标注结果如表 7、表 8 所示。

表 7 模型识别实例 1

Table 7 model identification example 1

	Only	France	and	Britain	backed	Fischler	‘s	proposal	.
人工标注	0	B-LOC	0	B-LOC	0	B-PER	0	0	0
BiLSTM-CRF	0	B-LOC	0	B-LOC	0	0	0	0	0
BiLSTM-BCRF	0	B-LOC	0	B-LOC	0	B-PER	0	0	0

表 8 模型识别实例 2

Table 8 model identification example 2

	Rare	Hendrix	song	draft	sells	for	almost	$	17000
人工标注	0	B-PER	0	0	0	0	0	0	0
BiLSTM-CRF	B-PER	I-PER	0	0	0	0	0	0	0
BiLSTM-BCRF	0	B-PER	0	0	0	0	0	0	0

在表 7 中可以看到该句子共有三个实体，原来的 BiLSTM-CRF 模型只实别出了两个实体，本文提出的 BiLSTM-BCRF 模型将句中包含的三个实体全部识别出来。表 8 中，该句子只有一个人名实体，但 BiLSTM-CRF 模型把句子中前两个单词错误地识别为一个人名实体，而本文提出的模型能将句中人名实体准确地识别出来。

5 总结

针对低资源领域标注语料较少的问题，本文提出了一种低资源场景下命名实体识别模型(BiLSTM-BCRF)。该模型通过在损失函数中融入伯努利分布，使得模型参数在低资源场景下也能得到较好拟合。同时将多层字符特征信息融合到模型中，提升模型处理罕见词的能力，使得模型在标注数据少量时也能拥有较好的识别性能，能更好的适应低资源命名实体识别任务。但该模型识别专有名词的性能还需继续提升，本文以后的工作会专注于提高模型识别专有名词的能力，同时模型的知识迁移和跨领域的性能提升也是本文以后的研究重点。

6 感谢

感谢审稿人提出的建设性意见，本文工作得到了国家自然科学基金(NO.61877031)的支持。

参考文献

[1]Hammerton J. Named entity recognition with long short-term memory[C]//Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003. 2003: 172-175.
[2]Lample G, Ballesteros M, Subramanian S, et al. Neural architectures for named entity recognition[J]. arXiv preprint arXiv:1603.01360, 2016.
[3]Huang Z, Xu W, Yu K. Bidirectional LSTM-CRF models for sequence tagging[J]. arXiv preprint arXiv:1508.01991, 2015.
[4]殷章志,李欣子,黄德根,李玖一.融合字词模型的中文命名实体识别研究[J].中文信息学报,2019,33(11):95-100+106.
[5]林广和,张绍武,林鸿飞.基于细粒度词表示的命名实体识别研究[J].中文信息学报,2018,32(11):62-71+78.
[6]Rathaparkhi A. A Maximum Entropy Part of Speech Tagger[J]. Proceedings of EMNLP'96, 1996.
[7]McCallum A, Freitag D, Pereira F C N. Maximum entropy Markov models for information extraction and segmentation[C]//Icml. 2000, 17(2000): 591-598.
[8]Lafferty J, McCallum A, Pereira F C N. Conditional random fields: Probabilistic models for segmenting and labeling sequence data[J]. 2001.
[9]Collobert R, Weston J, Bottou L, et al. Natural language processing (almost) from scratch[J]. Journal of machine learning research, 2011, 12(ARTICLE): 2493−2537.
[10]Chiu J P C, Nichols E. Named entity recognition with bidirectional LSTM-CNNs[J]. Transactions of the Association for Computational Linguistics, 2016, 4: 357-370.
[11] Ni J, Dinu G, Florian R. Weakly supervised cross-lingual named entity recognition via effective annotation and representation projection[J]. arXiv preprint arXiv:1707.02483, 2017.
[12] Mayhew S, Tsai C T, Roth D. Cheap translation for cross-lingual named entity recognition[C]//Proceedings of the 2017 conference on empirical methods in natural language processing. 2017: 2536-2545.
[13]Toolkit O S , BCK Học, MHN Ngữ, et al. Báo cáo khoa học: "Moses: Open Source Toolkit for Statistical Machine Translation"[J]. tailieu vn.
[14]Chen X, Awadallah A H, Hassan H, et al. Multi-source cross-lingual model transfer: Learning what to share[J]. arXiv preprint arXiv:1810.03552, 2018.
[15] Keung P, Lu Y, Bhardwaj V. Adversarial learning with contextual embeddings for zero-resource cross-lingual classification and NER[J]. arXiv preprint arXiv:1909.00153, 2019.
[16] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[J]. Advances in neural information processing systems, 2014, 27.
[17]Dai X, Adel H. An analysis of simple data augmentation for named entity recognition[J]. arXiv preprint arXiv:2010.11683, 2020.
[18] Chen J, Wang Z, Tian R, et al. Local additivity based data augmentation for semi-supervised NER[J]. arXiv preprint arXiv:2010.01677, 2020.
[19] Yang Y, Chen W, Li Z, et al. Distantly supervised NER with partial annotation learning and reinforcement learning[C]//Proceedings of the 27th International Conference on Computational Linguistics. 2018: 2159-2169.
[20]Tsuboi Y, Kashima H, Mori S, et al. Training conditional random fields using incomplete annotations[C]//Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008). 2008: 897-904.
[21] Feng J, Huang M, Zhao L, et al. Reinforcement learning for relation classification from noisy data[C]//Proceedings of the aaai conference on artificial intelligence. 2018, 32(1).
[22]Zhang T, Xia C, Philip S Y, et al. PDALN: Progressive Domain Adaptation over a Pre-trained Model for Low-Resource Cross-Domain Named Entity Recognition[C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. 2021: 5441-5451.
[23] Chen S, Pei Y, Ke Z, et al. Low-Resource Named Entity Recognition via the Pre-Training Model[J]. Symmetry, 2021, 13(5): 786.
[24]Pennington J, Socher R, Manning C D. Glove: Global vectors for word representation[C]//Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 2014: 1532-1543.
[25]Jie Z, Xie P, Lu W, et al. Better modeling of incomplete annotations for named entity recognition[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). 2019: 729-734.
[26]Sang E F, De Meulder F. Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition[J]. arXiv preprint cs/0306050, 2003.
[27]Li J, Sun Y, Johnson R J, et al. BioCreative V CDR task corpus: a resource for chemical disease relation extraction[J]. Database, 2016, 2016.
[28]Lin B Y , Lee D H , Shen M , et al. TriggerNER: Learning with Entity Triggers as Explanations for Named Entity Recognition[C]// arXiv. arXiv, 2020.

作者贡献声明：

钟茂生：设计模型思路、论文撰写和最终版本修订；

吴佳华：执行实验和论文修订；

所有作者声明不存在利益冲突关系

支撑数据：

支撑数据由作者自存储，202041600071@jxnu.edu.cn